# Subword complexity and finite characteristic numbers

Alina FIRICEL

Institut Camille Jordan, Université Lyon 1

CIRM, March 26, 2009

# The world of real numbers

The sequence of digits of the $b$-ary expansion of real numbers like:

$$\sqrt{2} := 1.0110101001\cdots \text{ or } \pi := 3.14159265358\cdots$$

is a source of difficult questions. Are there any simple “rules” or any evident patterns? Or is this “random”?

- The $b$-ary expansion of a rational number is eventually periodic. In particular, the structure of its sequence of digits is very simple.
- What about the real irrational algebraic numbers?
  - The sequence of digits of an irrational algebraic number cannot be generated by a finite automaton.
  - Conjecture: every irrational algebraic number is normal.
- What about the classical transcendental constants like $\pi$ or $e$?
- Many problems related to this topic are open!

# Formal power series with coefficients in a finite field

- Finite fields $\Rightarrow$ no more carry-over difficulties.
- There is a well-known analogy:

$$\begin{array}{ccc} \mathbb{Z} & & \mathbb{F}_p[T] \\ \cap & & \cap \\ \mathbb{Q} & \approx & \mathbb{F}_p(T) \\ \cap & & \cap \\ \mathbb{R} & & \mathbb{F}_p((1/T)) \\ \cap & & \cap \\ \mathbb{C} & & C \end{array}$$

- We replace the real number

$$\sum_{n \geq -k} \frac{a_n}{p^n}$$

by the formal power series

$$f(T) = \sum_{n \geq -k} \frac{a_n}{T^n} \in \mathbb{F}_p((1/T)).$$

# Objective

- Purpose: Study the subword complexity of formal power series with coefficients in a finite field in function of their arithmetic properties.

- What can we say about the Laurent series expansion of analogs of real numbers like $\sqrt{2}$, $\pi$, $\zeta(s)$, $e$?

- Study the effect of usual operations over formal power series. Closure properties of formal power series of "low" complexity (addition, multiplication, Hadamard product, derivative, Cartier operator,...).

# Complexity

Let $\mathbf{a} = a_0a_1a_2\cdots \in A^{\mathbb{N}}$. ($A$ finite alphabet)

- The subword complexity function is defined as follows:

$$p(\mathbf{a}, m) = \text{Card}\left\{(a_j, a_{j+1}, \ldots, a_{j+m-1}) \,:\, j \geq 0\right\}.$$

- **Example 1.** If $\mathbf{a} = aaa\cdots$ then $p(\mathbf{a}, m) = 1$ for every nonnegative integer $m$.
- **Example 2.** Champernowne's word: $\mathbf{c} = 01234567891011121314\cdots$ verifies $p(\mathbf{c}, m) = 10^m$.
- Remark: $1 \leq p(\mathbf{a}, m) \leq (Card\ A)^m$, for every integer $m \geq 0$.
- Let $f(T) = \sum a_n T^{-n} \in \mathbb{F}_p[[T^{-1}]]$. We define the complexity function of $f$:

$$p(f, m) := p(\mathbf{a}, m)$$

- If $\alpha$ is a real number whose $b$-ary expansion

$$\alpha = a_0.a_1a_2\cdots,$$

then the complexity of $\alpha$ in base $b$: $p(\alpha, b, m) := p(\mathbf{a}, m)$.

Theorem (Morse–Hedlund, 1938)

*$\boldsymbol{a}$ is eventually periodic if and only if $p(\boldsymbol{a}, m)$ is bounded. If not, the complexity function is strictly increasing. In particular,*

$$p(\boldsymbol{a}, m) \geq m + 1,$$

*for every nonnegative integer $m$.*

- $f \in \mathbb{F}_p(T) \iff p(f, m) = O(1)$.
- $\alpha \in \mathbb{Q} \iff p(\alpha, b, m) = O(1)$.
- $p(\pi, b, m) \geq m + 1$. Conjecture: $p(\pi, b, m) = b^m$.

# Algebraic formal power series

- Example: $f(T) = \sum_{n \geq 0} T^{-2^n} \in \mathbb{F}_2[[T^{-1}]]$. We have:

$$Tf^2(T) + Tf(T) + 1 = 0.$$

- The sequence of coefficients of $f$:

$$a_n = \begin{cases} 1 \text{ if } n \text{ is a power of } 2; \\ 0 \text{ otherwise} \end{cases}$$

is generated by a finite automaton.

J.-P. Allouche & J. Shallit
*Automatic Sequences: Theory, Applications, Generalizations*
Cambridge University Press, Cambridge, 2003.

- Moreover $p(f, m) = O(m)$.

# Algebraic formal power series

**Theorem (Christol, 1979)**

*Let* $f(T) = \sum_{n\geq 0} a_n (1/T)^n$ *be a formal power series with coefficients in* $\mathbb{F}_q$*. Then* $f$ *is algebraic over* $\mathbb{F}_q(T)$ *if, and only if, the sequence* $(a_n)_{n\geq 0}$ *is* $q$*-automatic.*

Cobham, 1972: If $\mathbf{a}$ is an automatic sequence then $p(\mathbf{a}, m) = O(m)$.

**Theorem**

*Let* $f \in \mathbb{F}_p((1/T))$ *algebraic over* $\mathbb{F}_p(T)$*. Then*

$$p(f, m) = O(m).$$

- *Real numbers:* If $\alpha$ is an irrational algebraic number and $b \geq 2$ then:

$$\lim_{m\to+\infty} \frac{p(\alpha, b, m)}{m} = +\infty.$$

B.Adamczewski & Y.Bugeaud
*On the complexity of algebraic numbers I. Expansions in integer bases*
Annals of Math. 165 (2007), 547–565.

# Carlitz' analogs

- Carlitz module over $\mathbb{F}_q[T] \Rightarrow$ an unique exponential $e_C(z)$ defined over $C$ by:

$$e_C(z) = z \prod_{a \in \mathbb{F}_q[T], a \neq 0} (1 - \frac{z}{a\widetilde{\Pi}_q})$$

  where

$$\widetilde{\Pi}_q = (-T)^{\frac{q}{q-1}} \prod_{j=1}^{\infty} \left(1 - \frac{1}{T^{q^j - 1}}\right)^{-1}.$$

- Remark: $\ker e_C = F_q[T]\widetilde{\Pi}_q$.

- Remark: $\widetilde{\Pi}_q$ is an analogue of $2i\pi$.
  Indeed, it is easy to verify that $\widetilde{\Pi}_q$ is a period of $e_C(z)$ (as $2i\pi$ is the period of $e(z)$).

- In order to obtain a formal power series associated to $\widetilde{\Pi}_q$ we take its real part:

$$\Pi_q = \prod_{j=1}^{\infty} \left(1 - \frac{1}{T^{q^j - 1}}\right)^{-1}.$$

# Complexity of $\frac{1}{\Pi_q}$

The power series expansion of

$$\Pi_q = \prod_{j=1}^{\infty} \left(1 - \frac{1}{T^{q^j - 1}}\right)^{-1} = \sum_{n \geq 0} p_n T^{-n},$$

$p_n =$ number of partitions of $n$ whose parts take values in $I = \{q^j - 1, j \geq 1\} \bmod p$.

BUT the power series expansion of:

$$\frac{1}{\Pi_q} = \sum_{n=0}^{\infty} a_n T^{-n}$$

$$a_n = \begin{cases} (-1)^{\operatorname{card} J} & \text{if } n \text{ can be written as } \sum_{j \in J} (q^j - 1); \\ 0 \quad \text{if} \quad n & \text{cannot be represented as a sum } \sum_{j \in J} (q^j - 1) \text{ for no finite set } J. \end{cases}$$


J.-P. Allouche
*Sur la transcendance de la série formelle* $\Pi$
J. Théor. Nombres Bordeaux 2 (1990), 103–117.

# Results

**Theorem**

*If* $q = 2$ *then* $p(\frac{1}{\Pi_2}, m) = \Theta(m^2)$*. More precisely:*

$$\frac{(m - \log m)(m - \log m + 1)}{2} \leq p(\frac{1}{\Pi_2}, m) \leq \frac{m^2}{2} + \frac{5m}{2}.$$

*If* $q \geq 3$ *then* $p(\frac{1}{\Pi_q}, m) = \Theta(m)$*. More precisely:*

$$p(\frac{1}{\Pi_q}, m) \leq 6qm.$$

**Corollary**

$\Pi_2$ *is transcendental over* $\mathbb{F}_2(T)$*.*

# Sketch of the proof for $q = 2$

Let us recall the first terms of the sequence:

| $n$ | $n$ | $a_n$ |
|---|---|---|
| 0 | 0 | 0 |
| 1 | $2^1 - 1$ | 1 |
| 2 | 2 | 0 |
| 3 | $2^2 - 1$ | 1 |
| 4 | $2^2 - 1 + 2 - 1$ | 1 |
| 5 | 5 | 0 |
| 6 | 6 | 0 |
| 7 | $2^3 - 1$ | 1 |
| 8 | $2^3 - 1 + 2 - 1$ | 1 |
| 9 | 9 | 0 |
| 10 | $2^3 - 1 + 2^2 - 1$ | 1 |
| 11 | $2^3 - 1 + 2^2 - 1 + 2 - 1$ | 1 |
| 12 | 12 | 0 |
| 13 | 13 | 0 |
| 14 | 14 | 0 |
| 15 | $2^4 - 1$ | 1 |

## Sketch of the proof for $q = 2$

- We denote, for $n \geq 1$, $w_n$ the subword of **a** defined as:

$$w_n = a_{2^n - 1} \cdots a_{2^{n+1} - 2}.$$

- Convention: $w_0 = 0$.
- Examples: $w_1 = 10$, $w_2 = 1100$, $w_3 = 11011000$ etc.
- Under these notations, the infinite word **a** may be written as:

$$\mathbf{a} = \underbrace{0}_{w_0}\underbrace{10}_{w_1}\underbrace{1100}_{w_2}\underbrace{11011000}_{w_3}\cdots = w_0 w_1 w_2 \cdots .$$

### Lemma

*For every $n \geq 2$ we have the relation: $w_n = 1 w_1 w_2 \cdots w_{n-1} 0$.*

- We denote $u_n$ the subword: $w_n = u_n 0$, for $n \geq 0$. For example $u_0$ is the empty word, $u_1 = 1$, $u_2 = 110$ etc.
- $|u_n| = 2^n - 1$

# Sketch of the proof for $q = 2$

For $m \in \mathbb{N}$ fix there exists an integer $n$ (greater or equal to 1):

$$2^{n-1} < m \leq 2^n.$$

**Goal:** calculate $p(\mathbf{a}, m)$

- Main idea: We denote $A_n = \{u_n^2 0^k, k \geq 1\}$. We prove that $\mathbf{a} \in A_n^{\mathbb{N}}$.
- Next step: All distinct words of length $m$ occur in the prefix of length $2^{n+1} - 1$ of $\mathbf{a}$, more precisely in:

$$u_n \underbrace{u_n 0}_{w_n} \underbrace{u_n u_n 00}_{w_{n+1}} \underbrace{u_n u_n 0 u_n u_n 000}_{w_{n+2}} \cdots w_m.$$

**Upper bound:** We look in all the words of the form: $u_n u_n$ (-> $2m$) and in all the overlaps $u_n 0^k u_n$ where $1 \leq k \leq m$ (-> $m - k + 1$).

**Lower bound:** It is sufficient to find

$$\frac{(m - \log m)(m - \log m + 1)}{2}$$

distinct words in $\mathbf{a}$.

# A vector space over $\mathbb{F}_p(T)$

The set of formal power series of polynomial complexity:

$$\mathcal{P} = \{f \in \mathbb{F}_p[[T^{-1}]], \text{there exists } K \text{ such that } \; p(f,m) = O(m^K)\}$$

Remark: Algebraic power series belong to $\mathcal{P}$; $\frac{1}{\Pi_q} \in \mathcal{P}$.

**Theorem**

*$\mathcal{P}$ is a vector space over $\mathbb{F}_p(T)$.*

Remark:

- Moreover $\mathcal{P}$ is closed under Hadamard product, (formal) derivative, Cartier operator...
- The same properties are satisfied by the set of formal power series of entropy 0.
- **Consequence:** This leads to a criterion of linear independence.

Subword complexity $\leftrightarrow$ Space, time complexity


D.Thakur & R. Beals
*Computational classification of numbers and algebraic properties*
International Mathematics Research Notices, 15 (1998), 799–818.

## Conclusion and perspectives

- Study other closure properties of formal power series of low complexity, such as multiplication or inverse. Stability under these operations could imply in particular algebraic independence over $\mathbb{F}_q(T)$.

- There is some particular cases of formal power series belonging to $\mathcal{P}$ stable by multiplication (some lacunary power formal series, automatic series). But see for example Jacobi theta function:

$$\begin{aligned} \theta_3(T) &= 1 + 2\sum_{n\geq 1} T^{-n^2} \in \mathbb{F}_q((1/T)),\ q \geq 3 \\ \theta_3^2(T) &= \sum_{n\geq 1} r_2(n)T^{-n} \text{ where } r_2(n) = 4(d_1(n) - d_3(n)) \bmod q \end{aligned}.$$

  It is not difficult to prove that $\theta_3(T) \in \mathcal{P}$. But what about $\theta_3^2(T)$ ?

  *Difficulty:* Study the complexity of $\theta_3^2(T) \Rightarrow$ study additive properties of the multiplicative sequence $(r_2(n))_n$.

- Place other well-known transcendental finite characteristic numbers such as $e$ or Carlitz $\zeta$ values in the computational hierarchy.

- Formal power series of low complexity $\Rightarrow$ diophantine properties. (linear complexity $\Rightarrow$ irrationality measures)